



RCRAInfo V2 UNIVERSE CALCULATIONS AND REPORTING

Universe calculations, both how they are done and how the data is stored, are a vital part of RCRAInfo. The universe calculations establish basic programmatic determinations for a site in a consistent manner for all users of the data. RCRAInfo Version 2 (V2) continues to enhance the usefulness of the universe calculations present in RCRAInfo Version 1 (V1) while incorporating changes required by the WIN/INFORMED: Universe Identification, Waste Activity Monitoring, Program Area Analysis (PAA).

This document is divided into two major sections. The first section explains the basic design concepts and programmatic interpretations established by the V2 Design Team. The team is composed of representatives from the States, Regions, and EPA HQ. The second section communicates basic technical information concerning the Universe Calculations that is necessary for proper understanding of the universe calculations.

I. Design Concepts and Programmatic Interpretations

1. Sources of Information

The requirements for Universes existing in RCRAInfo V1 are documented in the RCRAInfo On-line HELP utility.

The requirements for universes added to RCRAInfo V2 are found in the WIN/INFORMED: Universe Identification, Waste Activity Monitoring, Program Area Analysis: Final Report available at <http://www.epa.gov/epaoswer/hazwaste/data/win/r00-004.pdf>. Of particular interest is Appendix V: TSDf Universe Definitions.

2. Basic Design

While designing the universe calculations for V2, the Design Team balanced a number of different design criteria. The goal was to develop universe calculations that were timely, flexible, easy to use, easy to maintain, and supported numerous different data reporting uses. The Design team developed a data structure containing three tables: Hreport_univ, Huniverse_detail, and Huniverse_just.

Hreport_univ is the main table that stores the results of the universe calculation. Hreport_univ is the largest and most comprehensive of the tables, and is designed to be the main source of information when reports containing universes are created. In addition to storing the results of the calculation, this table also stores information commonly used in most RCRAInfo reports. The type of data stored in addition to the universes is site level handler information. This structure allows for easy report design and development.

Huniverse_detail is a smaller table than Hreport_univ and contains detailed breakdowns of certain universes. Huniverse_just was designed to support data quality by recording why handlers were placed in each universe. The intent of Huniverse_just is to allow system users to easily determine what specific piece of information is placing specific handlers in each, and every, universe.

In response to our experience with RCRAInfo V1, the Hreport_univ table was designed for easier reporting by placing information for each universe for each handler (regardless of whether they were in the universe or not) in one row, which is different than the design for Huniverse (which only included information when a handler was included in a universe).

3. What rows are present in Hreport_univ

Hreport_univ contains one record for each combination of Handler ID and Activity Location present in the Handler tables and the Compliance Monitoring and Enforcement (CME) tables. Technically this means that there exists in Hreport_univ one row for each unique combination of Handler_ID and Activity_Location present in the Hhandler2, Cenforcement, Cevaluation and Cviolation tables. The presence of the CME tables in this list was necessary because RCRAInfo was designed to allow users to create CME records without having to enter corresponding Handler records (this was primarily done to allow for out-of-state transporter inspections). Hreport_univ records that do not have corresponding Hhandler2 records can be easily identified because the field Source_type in Hreport_univ is equal to 'O' (for Out of State). (Note: Source_type 'O' is only valid in the Hreport_univ table). **In order to simplify report development, Handler type information (e.g., name, address, contact) will be populated for the Source_type 'O' records from the home State record for the ID (the home State is the State that matches the first two characters of the EPA ID). The Design Team is working through the details of exactly which data elements will be populated in this manner and this information will be distributed as soon as it is determined.**

Note: The exception to this rule is Handlers who only have Biennial Report source data from the 1999 Biennial Report cycle and before (that is, in the Hhandler2 table, all records for a single ID have a source_type equal to 'R' and report_cycle <= 1999) are excluded from the Hreport_univ table¹. This exclusion applies regardless of whether the site has any other permitting, corrective action, or compliance monitoring and enforcement data. The reason for this exclusion is this data is highly "suspect", since all sites are required to have a notification (or implementer inspection) record. This changed with the 2001 Biennial Report cycle, where a site could file the new Site ID Form to satisfy their biennial reporting requirements, as well as for other purposes, such as updating their Notification data. The Design Team plans to develop a new data quality report to help implementers identify these sites, so that proper notification forms can be entered into the system.

¹ For the purposes of allowing RCRAInfo system development work to proceed before the 2001 Biennial Report data has been loaded into RCRAInfo, the universe calculation programs will be set to temporarily allow the 1999 Biennial Report data. Once the 2001 Biennial Report data has been loaded (before the end of 2002), the universe calculations will be re-run and the 1999 data will be excluded.

4. Merging Waste Activity Monitoring and General RCRA Data

While reading thru the various PAA reports, the Design Team noticed a number of different themes. One theme in particular was the need to easily obtain current information for handlers in the database, and to have that data available for easy reporting. Currently, RCRAInfo supports three main sources of handler information. The first source is the initial handler notification. The second source is EPA or State inspections of a handler. The third source of information is from Biennial Report data submitted by a handler. Data from each of these sources comes in at different times for different handlers. The Biennial Report data comes in every two years, but is only required for Large Quantity Generators (LQGs) and Treatment, Storage, and Disposal Facilities (TSDFs). Notification data is only required to be updated under certain conditions, except for those implementers that specifically require re-notification on a regular basis. The information from one data flow can contradict information from another data flow. Also, many report writers have expressed a need to have a single source of data for easy report writing.

Historically this information was kept in separate systems, complicating reporting and data analysis. Notification and inspection (General RCRA) data was kept in RCRIS, while Biennial Report (Waste Activity Monitoring) data was kept in BRS. Even in RCRAInfo V1 the data was kept in separate tables, even though the data was in the same database. Data reporting was complicated because the report writer needed to understand the various systems and data flows in order to produce correct reports.

Another theme of the PAA reports was to better integrate data across the various information sources. In order to support the PAA recommendation on data integration, the Design Team re-designed the Handler module data structure to accommodate data from both Waste Activity Monitoring and General RCRA data sources. This had the intentional effect of, for the first time, being able to compare, easily and directly, handler data from all data sources. This also had the intentional effect of forcing changes to the existing universe calculations, where these universe calculations involved Handler data. The universe calculations that were affected were the generator status and transporter universes (as well as affecting the universal waste and used oil universes new in RCRAInfo V2). (Note: for ease of discussion these universes will be collectively referred to as the Handler universes.)

In RCRAInfo V1 (and RCRIS before) the Handler universes were determined using two basic algorithms. The first was to calculate the universes based on a hierarchy of information available for each site. The hierarchy was limited to those records contained in the RCRIS database, and consisted of Notification records as well as EPA and State inspection records. Other source records (including, for example, Part A records) were ignored in the calculation. This algorithm was changed in the later stages of the RCRIS system to a hierarchy based on date, where the most recent record was used to calculate the Handler universes, and the previous records were ignored. (In cases where there were multiple records that corresponded to the most recent date a source based hierarchy was used as a secondary step). Once again, certain data sources like Part A data were ignored. This same algorithm was used in RCRAInfo V1.

While working on RCRAInfo V2, the Design Team confronted the question of how the universe calculations should be affected by the merging of the Handler data. A variety of options were proposed and debated. The Design Team consulted with the PAA Team to determine which option was most in keeping with the PAA Team's guidance and directions. The selected option, presented below, is the option that the team feels is most in keeping with the PAA guidance, is consistent with the historical data, and is supported by the new data streams coming into the system.

The decision by the Design Team was to calculate the Handler universes based on which data source is the most current data source, with the exception of Biennial Report source records from 1999 and before (that is, Hhandler2 records with source_type = 'R' and report_cycle <= 1999 are excluded from the calculation)². The implicit assumption is that the most current data source contains the best data, that is, the data that is most indicative of the real status of the handler. Data from certain Biennial Report cycles is excluded from this calculation because the data collection instrument used to collect the data was substantially different than the existing notification form, did not ask the same questions as the notification form, and collected substantially less data than the notification form. The differences between the Biennial Report form and the notification form were eliminated in the 2001 Biennial Report cycle, which is why the 2001 (and beyond) Biennial Report data is included in the calculation.

(NOTE: Similar to the second algorithm used in RCRIS, there is a hierarchy used in cases where there is more than one most current data source. In these cases, the following source hierarchy is used to determine the dominate data record: I, N, B, R, A, T, E, where 'I' corresponds to implementer source records, 'N' corresponds to notification source records, 'B' corresponds to records submitted both as a notification and as a Biennial Report record, 'R' corresponds to Biennial Report only records, 'A' corresponds to Part A applications, 'T' corresponds to temporary source records, and 'E' corresponds to emergency source records.)

The Design Team realized, of course, that implementation of the new algorithm, combined with the conversion of data from V1 to V2, will change the Handler universes for some sites in the database. For example, there are sites in the database whose most recent record is a Part A application but has a much older notification record. According to the old algorithm the Part A application record was ignored, and the handler universes were based on the notification record. The new algorithm would change that by basing the handler universes on the Part A record, because it is the most recent record.

² For the purposes of allowing RCRAInfo system development work to proceed before the 2001 Biennial Report data has been loaded into RCRAInfo, the universe calculation programs will be set to temporarily allow the 1999 Biennial Report data. Once the 2001 Biennial Report data has been loaded (before the end of 2002), the universe calculations will be re-run and the 1999 data will be excluded.

The Design Team's choice of algorithm supports this intent of the PAA, and still allows implementers to use the flexibility of the different source records in determining universe calculations. If an implementer disagrees with the outcome of the universe calculation they can easily locate the data record where the information came from and enter a new record for the site. This also supports future PAA recommendations concerning periodic re-notification and electronic re-notification by handlers.

II. Technical Information

1. Table versus View

During the initial stages of V2 the development team attempted to design the Hreport_univ table as a series of Oracle views. Oracle views have the advantage that they are updated whenever the database is changed, and do not rely on a reoccurring batch job to update the view. Unfortunately, preliminary testing by the Design Team indicated that the speed of the views was far too slow to allow for adequate report performance and as such the views were changed to tables.

2. Schedule

All fields in the universe calculation tables are updated every three (3) hours by an automatic Oracle job. Users can check when the last time the Universe Calculation were run by selecting the batch universe calculation results from the system administration menu. At present, this is the only way that universe calculations are updated. In the future, this may be changed to allow users to run universe calculations on selected sites, subject to the number of users requesting this feature and the programming expense to establish the feature. When necessary, all the universe calculation tables can be recreated "from scratch", however it is not known how often, or on what schedule, this will be performed.

3. Huniverse_detail

Huniverse_detail was designed to support detailed reporting on certain universes. The universes presented in Huniverse_detail are also present in Hreport_univ, however the format of the presentation is different. In Hreport_univ the universes are represented by fields that are several characters long (for example, the permit workload universe in Hreport_univ is represented by one field that contains 5 characters). In Huniverse_detail the 5 character field is broken out into 5 one character fields. The intended use of these tables resulted in the two different designs. Hreport_univ was designed to allow for easy reporting of the entire universe at one time. Huniverse_detail was designed to allow selecting each one character component of universes that have multiple character universe values.

For this reason, there are intentionally many fewer rows in Huniverse_detail than in Hreport_univ. While Hreport_univ contains a record regardless of whether a handler is in a universe, Huniverse_detail contains a record only if the handler is part of one of the universes represented in the table. This policy save disk space, as well as dramatically decreasing the amount of time

necessary when searching thru the huniverse_detail table. Specifically, a record in Huniverse_detail will be created only for those records that contain at least one 'Y' in the Hreport_univ used_oil field (e.g. 'NNNNNNY') or at least one 'L' 'I' 'B' 'S' or 'T' in the Hreport_univ closwrkld, pcwrkld, full_enforcement, permprog, permwrkld fields.

4. Huniverse_just

Huniverse_just was designed to allow users easy access to information about why a handler was included in each, and every, universe. Each row of the table provides details on one specific universe for one specific activity location for one specific handler. Huniverse_just will contain multiple rows of data if a handler is included in a specific universe based on more than set of data. In this case, each row will detail one specific data set. This allows users to comprehensively see all the reasons why a handler is included in specific universes.

The primary data field of interest in the Huniverse_just table that is of interest to users is the Justification field. This field records the specific reason why a handler is included in a specific universe. The field Universe details which universe the Justification field is detailing. The exact content of the Justification field will vary according to the universe being detailed. The field Event_Date in Huniverse_just is designed to allow for users to sort the data. The exact value that is placed into Event_Date varies according to the universe.

Users may access the Huniverse_just table in two different manners. The RCRAInfo data entry screens are being modified to display information from a specific handler when the "Universe" button on the main data entry screen is chosen. Alternatively, users may also do direct SQL queries against the table and create their own reports.

5. Miscellaneous Data Elements

The Design Team is still in the process of deliberating on how to populate those data elements present in the Hreport_univ table that are not present in the file specifications for the 2001 Biennial Report data files. This includes fields like the state_district field. RCRAInfo does not allow implementers to update Biennial Report data records (other than thru the load process) and the load process does not allow certain fields like state_district to be provided. The results of these deliberations will be communicated once the discussions are finished.

Please refer to newly updated structure chart for the Hreport_Univ Table for more details. The structure chart is located on the RCRAInfo Partnership Website: <http://www.epa.gov/oswfiles/rcrainfo/>